Bound Docking by Optimization of Electrostatic Interactions in Protein Complexes

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of

Master of Technology

by

Preeti Kumari
Roll No:- Y3118007

to the

DEPARTMENT OF BIOLOGICAL SCIENCES AND BIOENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY KANPUR, INDIA
MAY, 2005

Dated: 3 MAY, 2005.

CERTIFICATE

This is to certify that the work under the thesis titled **"Bound Docking by Optimization of Electrostatic Interaction in Protein Complexes"** by **Preeti Kumari (Roll No. Y3118007)** has been carried out under our supervision and this work has not been submitted elsewhere for a degree.

Bhaskar Dasgupta

Dr. Bhaskar Dasgupta
Department of Mechanical Engineering
Indian Institute of Technology
Kanpur
India.

(D.S. KATT

for Dr. Balaji Prakash
Department of Biological Sciences
and Bioengineering
Indian Institute of Technology
Kanpur
India.

Dedicated to

My Beloved Parents

Acknowledgment

I would like to take this opportunity to express my deep sense of gratitude to my thesis supervisor and teacher **Dr. Bhaskar Dasgupta** for allowing me to take up this challenging work under his supervision. I would like to thank him for his faith in me as a *student* to learn new things and for giving his valuable timely advices. Finally, I would like to thank him for his invaluable guidance and relentless encouragement.

I would also like to thank my thesis advisor **Dr. Balaji Prakash**, for encouraging me to take up this challenging work outside my deparment. I would like to thank him for his constant faith in me to do something different and good for my thesis. I would also like to thank the head of my department **Prof. Pradip Sinha** for allowing me to choose a thesis topic to be worked out in other deparment. I would also like to thank **Dr. R. Sankararamakrishnan** for his advices and encouragement at times of need. I would like to express my gratitudes to **Dr. Savitha Govardhan, Dr. Sumit Basu** and **Dr. Nandini Gupta** for their kind concern.

I would especially like to thank my friend *dhiraj* for enriching me by sharing his professional and emotional experiences and encouraging me to meet the challenge of learning new things for my thesis. I would also like to convey my gratitudes to my friend and labmate *hari* for giving me his valuable advices at times required for my thesis. I would also like to thank my friends *patelji, ranja, nikks, udit and negi* for sharing their love and warmth and providing me with a friendly atmosphere and cherishable moments at iitk.

I would like to thank my parents for enabling me to reach to this stage in my life, it is always their blessings and encouragement which gives me the courage to face every new challenge of my life with equal fervor and enthusiasm.

And above all, I thank the Almighty for enabling me to achieve higher goals.

Preeti Kumari
I.I.T. Kanpur
3 MAY, 2005.

Abstract

The algorithm presented in the thesis, aims at reconstructing a protein-inhibitor (protein) complex from the unbound protein and inhibitor structures, by searching for a minimal energy conformation on the basis of electrostatic interaction. The minimal energy conformation has been obtained by the optimization of the electrostatic interaction. For calculating the electrostatic interaction between the protein and inhibitor at the binding site, point to point Coulombic interaction method has been used, based on the continuum dielectric solvation model. The algorithm has been tested on ten different protein-inhibitor complexes and from the results obtained it can be concluded that, the role played by electrostatic interaction at the binding site between the protein-inhibitor complexes is case dependent, it plays a critical and important role when both the interacting surfaces have the presence and distribution of opposite potential surface patches *i.e.* the presence of oppositely charged side chains at the interface. This algorithm can be used for the purpose of secondary screening of the candidate solutions obtained by initial screening done on the basis of geometric complementarity to improve the ranks of nearly correct solutions, but the degree of success in improving the rank will be case dependent *i.e.* depending upon the presence of different potential patches at the interacting surfaces.

Contents

List of Figures

LIST OF PROTEIN-INHIBITOR COMPLEXES USED

Bound Complex	Unbound Protein/Inhibitor
(PDB Codes)	(PDB Codes)
1LDT	1EPT/1LDT
1AVW	1EPT/1AVU
1BRS	1BNI/1BTA
1BVN	1PIF/2AIT
1CHO	5CHA/1OVO
2PTC	2PTN/4PTI
1FSS	2ACE/1FSC
1SMF	2PTN/1PI2
2SEC	1SCD/1TEC
1TEC	1THM/2SEC

Chapter 1

Introduction

1.1 Structure and Function of Proteins

1.1.1 Basic Building Blocks: Amino Acids

Amino acids are the basic building blocks of proteins. There are 20 different amino acids found in all proteins. All the twenty amino acids have in common a central carbon atom (C_α) to which are attached a hydrogen atom, an amino group (NH_2) and a carboxyl group (COOH) as shown in the Fig 1.1. One amino acid is different from the other with respect to the side chain (R) attached to the C_α through its fourth valence, as there are 20 different types of side chains specified by the genetic code and thus, 20 different amino acids. The amino acids are abbreviated with both a three letter and one letter codes,

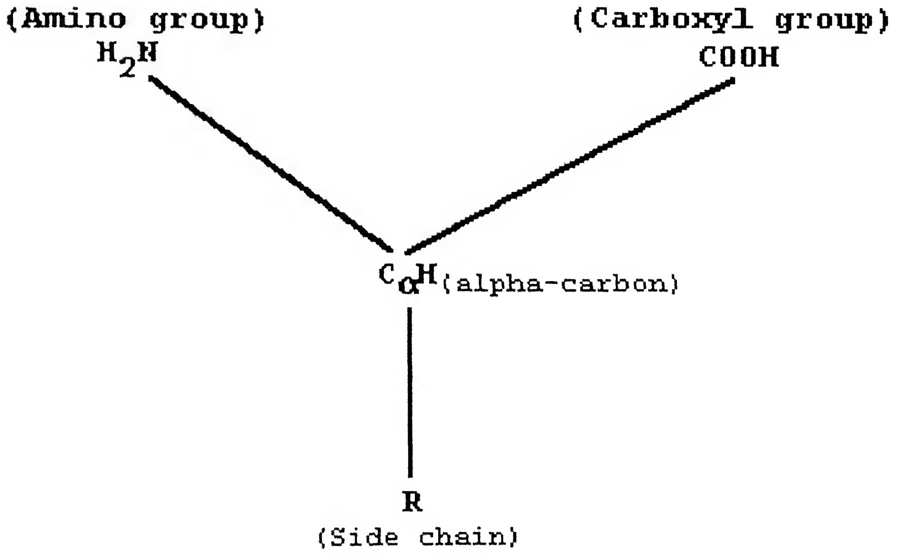


Figure 1.1: Amino acid structure

they have been listed along with their chemical structures in the Appendix B. The amino acids are usually divided into three different classes depending on the chemical nature of their side chains. The first class comprises of those with *non-polar or hydrophobic side chains* : ALA, VAL, LEU, ILE, PHE, PRO and MET. The second class is of four *charged residues* : ASP, GLU, LYS and ARG.The third class comprises of those with *polar side chains* : SER, THR, CYS, ASN, GLN, HIS, TYR and TRP. The amino acid GLY has only a hydrogen atom as a side chain and is the simplest amino acid and has special properties which can be used in protein structure determination. Due to the chiral C_α atom in amino acids(except GLY), they can be present in two forms the L-form and the D-form, all the amino acids that occur in proteins are in L-form. Amino acids are joined end to end during protein synthesis by **peptide bonds** which is formed when the carboxyl (COOH) group of one amino acid condenses with the amino group (NH_2) group of the next with the elimination of water as seen in Fig 1.2.

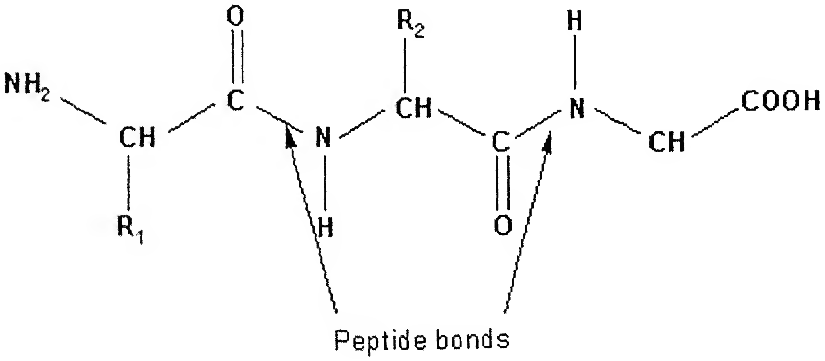


Figure 1.2: Peptide bond formation in amino acids

1.1.2 Primary, Secondary, Tertiary and Quaternary Structure

The **primary structure** of a segment of a polypeptide chain or of a protein is the amino-acid sequence of the polypeptide chain. The C_α atoms of amino acids form the main-chain atoms of the protein backbone to which are attached the side chains.The sequence and properties of side chains determine all that is unique about a particular protein, including its biological function and its specific three-dimensional structure.

The **secondary structure** of a segment of polypeptide chain is the local spatial arrangement of its main-chain atoms without regard to the conformation of its side chains or to its relationship with other segments. There are three common secondary structures in proteins, namely *alpha helices, beta sheets and turns*. An extremely useful device for studying protein conformation is the Ramachandran plot which plots ϕ and ψ (Fig 1.3). The values of ϕ and ψ that are possible are constrained geometrically due to steric clashes between neighboring side chains. The values of ϕ and ψ can be plotted on a two-dimensional map of the $\phi - \psi$ plane which shows allowed and disallowed regions. Regular secondary structure conformations in segments of a

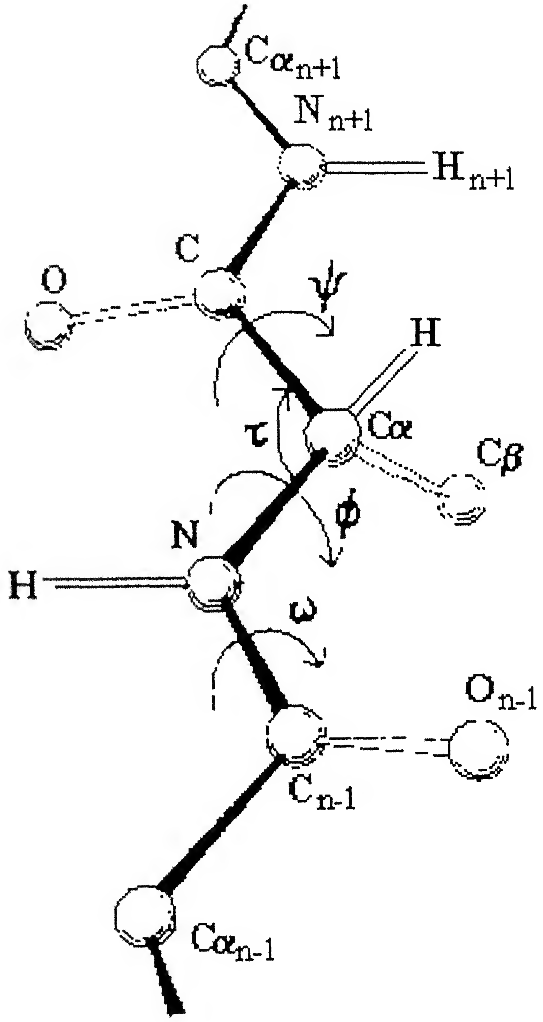


Figure 1.3: $\phi - \psi$ and ω angles in the peptide chain

polypeptide chain occur when all the ϕ torsion angles in that polypeptide segment are equal to each other, and all the ψ torsion angles are equal. The alpha-helix and beta sheet structure conformations for polypeptide chains are generally the most thermodynamically stable of the regular secondary structures. However, particular amino acid sequences of a primary structure in a protein may support regular conformations of the polypeptide chain other than alpha-helical or beta-structure. Thus, whereas alpha-helical or beta-structure are found most commonly, the actual conformation is dependent on the particular physical properties generated by the sequence present in the polypeptide chain and the solution (surrounding) conditions in which the protein is present. In addition, in most proteins there are significant regions of disordered structure in which the ϕ and ψ angles are not repetitive, these are called *loop- regions.*

The alpha helices are found when a stretch of consecutive residues all have the (ϕ, ψ) angle pair approximately -60^o and -50^o. The α helix has 3.6 residues per turn with hydrogen bonds between C=O of residue n and NH of residue $n+4$. Thus, all NH and CO groups are held with H-bonds except the first NH groups and the last CO group, which are at the ends of the α helix, making their ends polar. Some variants of alpha helix are 3_{10} and pi helices but, they are present rarely in proteins. The α helices can vary in length from four or five to over forty residues.

The second major structural element found in proteins are the β sheets. These β sheets in turn are made up of structural units called β strands, which are 5 to 10 residues long, with (ϕ, ψ) angles ranging from -120^o to 120^o. Beta sheets are made up from a combination of several regions of the polypeptide chain, unlike α helices, which are built from a single continuous region. The β strands are aligned adjacent to each other such that hydrogen bonds can form between C=O groups of one strand and NH groups on an adjacent β strand. The β sheets formed from several β strands give a *pleated* appearance due to C_α atoms successively a little above and below the plane of the β sheet. When the β strands are arranged parallel to each other then the sheet is described as *parallel* while if they are arranged anti parallel the sheet is called *anti parallel.*

Most proteins are built up of certain combinations of secondary structural elements, α helices and β sheets which are connected by *loop regions* of various lengths and irregular shapes, such combination of structure is called the **tertiary structure** of the protein. A combination of the secondary structural elements in the tertiary structure tend to form a stable hydrophobic

core of the molecule. The loop regions are exposed to the solvent and are rich in charged and polar side chains, these regions in addition to forming the connection between the secondary structural elements also frequently participate in forming binding sites and enzyme active sites.
When a protein is made of more than one polypeptide chain, the structure formed by the assembly and spatial arrangement of all the polypetide chains is called the **quaternary structure** of the protein, the polypeptide chains can be the same or different. Such proteins are called *multi-domain* proteins.

1.1.3 Forces Stabilizing the Protein Structure

Proteins perform many of the functional roles necessary for living systems. The function of most proteins is based upon the unique conformation that their polypeptide chain adopts in solution (*i.e.* the *native* or *folded* conformation). The folded conformation is determined by the primary sequence and the interaction of amino acid side chains with the solvent (surrounding medium). The pH, ionic composition and concentration, and the solvent dielectric, can influence the electrostatic interactions that stabilize the folded conformation. *Following are some non-covalent interactions stabilizing protein structures.*
Hydrogen Bonds : A hydrogen bond is formed when a hydrogen atom with a large positive partial charge interacts with an atom with a large negative partial charge. The opposite charges attract each other and the hydrogen atom which is covalently bound to the hydrogen bond *donor* atom comes very close to the hydrogen bond *acceptor* atom with its lone pairs. In general, the two partial charges (positive and negative) are part of dipoles, which causes the positive hydrogen to be positioned between two electronegative atoms as seen in Fig 1.4. In protein molecules polar and charged side chains of amino acids take part in hydrogen bonding. Hydrogen bonding plays an important role in forming the secondary and tertiary structure of protein molecules and in protein-protein association.

Hydrophobic Interactions : Hydrophobic interactions are the most important non covalent force that cause the linear polypeptide to fold into a compact structure. But, it is not the interactions between side chains of hydrophobic amino acids (which is mainly van der Waals interaction) that

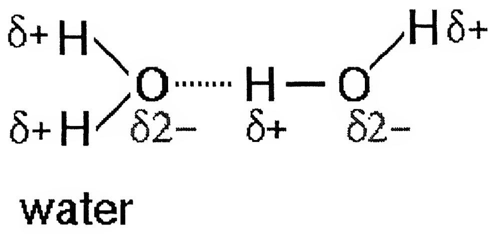


Figure 1.4: Hydrogen Bonding

induce the strong interaction, but the increase in entropy gained by the removal of hydrophobic surface from the aqueous (polar) environment. The aggregation of the hydrophobic surfaces (consisting of non-polar residues) forms the tightly packed core of a protein, as it is a primary driving force for protein folding which causes the removal of non-polar side chains from polar solvent exposure. Some non-polar groups or hydrophobic patches are found on the surface of the protein molecules as determinant of preferred sites of molecular associations [10].

Electrostatic Interactions : Proteins bear many polar and charged side chains mostly on their surface, playing an important role in protein-protein interaction. In one study [3] it has been found that the electrostatic energy of interaction between protein complexes strongly correlates with the rate of association. At the protein interface there is optimization of interaction between charged and polar groups, which in turn produces interactions that are stabilizing, highly directional and distant-dependent, allowing the significant specificity that is characteristic of many recognition processes involving biological macromolecules. The electrostatic interaction can be described by Coulomb's interaction method in which the energy is related to the inverse of the distance between the interacting charges, this method is used in the algorithm presented in the thesis for the calculation of electrostatic energy.

Van der Waals Interactions : Important contributions to protein stability are given by the London dispersion forces (attractive and distant) and electron shell repulsion (repulsive and close). The attractive component is due to the induction of dipoles in the electron cloud of neighboring atoms, coupling of the dipoles, leads to attractive forces. The repulsive component is due to the sterical hindrance when neighboring atoms start to have overlap of the electron clouds. The attractive (distant) and the repulsive (close) components are usually taken together and described by the Lennard-Jones potential as follows :

$$E = E_m[-(R_m/R_{ij})^{12} + 2(R_m/R_{ij})^6] \tag{1.1}$$

The second term shows weak attraction at long distance (power of 6 term) and the first term shows strong repulsion at very close distance (power of 12 term). At *bonding* distances there is an energy well at R_m where the van der Waals energy is at a minimum (E_m). The repulsion energy is small for the closest contact distance R_{vdw} which is the sum of the so-called van der Waals radii for the two atoms.

1.1.4 Mode of Protein-Protein/Ligand Interactions

Protein-protein association involves the specific (at times non-specific) complementary recognition of two macromolecules to form a stable assembly. The formation and stabilization of the complex involves various non-covalent interactions occurring at the interface namely, electrostatic interaction, hydrogen bonding and presence of hydrophobic patches at the surface [10] are some of the important driving forces for the formation and stabilization of the complex. Formation of the complex reduces the net charge on the molecules, through interaction with other oppositely charged side chains present on the surface of the molecules and buries the exposed non-polar side chains present at the surface.

Following are the two hypothesis suggested for the mode of interactions between various protein (like enzymes) and its substrate molecule.

Lock and Key Hypothesis

An enzyme (a protein molecule) is globular and very large but only a small part of it, the active site, is involved in any reaction. When the shape of the active site matches with the that of the substrate molecule, the substrate molecule fits into the active site and is held there until the reaction is completed. The product is then released and the enzyme is once again ready to take part in another reaction. This is known as the lock and key hypothesis. The active site has a distinct shape, rather like a lock. Just as only the right *key* will fit a lock, so only the right substrate has the right shape to fit into the active site.

Induced-fit Hypothesis

The lock and key hypothesis, does not explain the interaction completely and efficiently as in that case some small molecules like water may enter and interfere in the reaction, moreover this hypothesis does not add flexibility to the binding site, so, later a refinement of this hypothesis was suggested *i.e.* the induced-fit hypothesis. According to this hypothesis [30], the substrate (ligand) does not simply bind with the active site. It has to bring about changes to the shape of the active site to activate the enzyme (catalytic protein) and make the reaction possible. So small molecules may enter the active site, but they cannot induce the changes in shape to make the enzyme active. The hypothesis suggests that when the enzyme's active site comes into contact with the right substrate, the active site slightly changes or moulds itself around the substrate for an effective fit. This shape adjustment triggers catalysis (reaction) and helps to explain why certain enzymes only catalyse specific reactions.

1.1.5 Summary

Proteins are macromolecules, made up of one or more polypeptide chains, which are in turn made up of 20 different kinds of amino acids. A typical protein contains 200-300 amino acids, but some can be much smaller which are called *peptides*. Proteins play an important role in the fundamental processes of the cell. Their function is determined by their primary sequence which in turn determines their structure. The α helix and β sheets are the most stable and commonly found secondary structural elements found in proteins. The stability of protein structure and its interaction with other proteins depends on many non-covalent interactions such as electrostatic interactions, hydrogen bonding, hydrophobic interactions and Van-der Waals interactions. The binding site of proteins have considerable flexibility as suggested by the induced-fit hypothesis and observed in various experimental studies.

1.2 Protein-Ligand Docking

1.2.1 Literature Review

Definition and Aim of Protein-ligand Docking

Protein-ligand docking can be defined as *For two given biological molecules determine whether they interact and if they interact then determine the orientation of their maximal interaction while minimizing the energy of the complex [1].* Here, the term ligand means either a protein molecule or a chemical agent used as in drugs. Aim or goal of protein-ligand docking can be defined as *To be able to search a database of molecular structures and retrieve all molecules that can interact with the query structure.*

Docking basically is of two types *Bound Docking* and *Unbound Docking* [1]. Bound docking deals with computational schemes that try to regenerate a complex from the bound structures of the protein and ligand. Thus, in this case the binding site is priorly known, such structures are mostly obtained from co-crystallized structures. While, unbound docking deals with those computational schemes that try to regenerate a complex from the unbound structures of the protein and ligand. So, this is the more difficult part of docking as here the a prior knowledge of binding sites is not available [1].

Different phases of Docking

The process of docking can be divided into three phases namely :

1. **Preprocessing Phase:** This phase deals with the mathematical representation of the system or mapping the three-dimensional surface of the receptor and ligand. Surface representation is done mostly by its geometric representation. Most common method used for this is Connoly surface representation [13]. Connoly surface consists of the part of the Van-der Waals surface of the atoms *i.e.* accessible to a probe sphere (contact surface) connected by a network of convex, concave and saddle shape surfaces that smooths or rolls over the crevices and pits between the atoms. Its a method to describe surface on the basis of sparse critical points. A surface normal at each point is generated, then the need is to detect a pair of critical points in both molecules that share the same internal distance and if superimposed, have opposing

surface normals.

The other commonly used method for surface representation is the grid method [2] *i.e.* representing the 3D surface of the protein on to a fine grid and assigning different scores for the points falling on the surface, in space and penalty for those inter penetrating. But, this penalty should be decided carefully as it should be neither too high nor too low, so that certain flexilibility is allowed at the interface.

2. **Recognition Phase:** Its the most important and critical phase of docking which involves recovering candidate ligands from the database generated in the pre-processing phase, matching the receptor/protein's surface patches and rank the candidates on the basis of the scores obtained [1].

3. **Post-processing Phase:** This phase deals with filtering out the best candidates out of the top ranking candidates obtained in the recognition phase. For this, electrostatic interactions, solvation energy and other kinds of interactions occurring at the interface can be taken into consideration to be used as a criteria for filtering the minimal energy state candidates out of the best [1].

Scoring Functions

Scoring Functions are used to score the candidates and rank them on the basis of scores obtained. Thus, scoring helps in detecting correct solutions with low ranks and those having minimum rmsd deviations from the crystal complex [1]. Based on the scoring functions used by an algorithm docking can be further divided into two types *i.e.* Geometric Docking and Integrated Docking algorithms [2]. Former kind of docking takes into consideration only the shape complementarity, ignoring any other kind of interactions such as electrostatic interactions occurring at the surface. While, the latter kind of docking also use some of the energy functions such as electrostatic interactions [12], solvation energy, H-bonding etc [5]. occurring at the interface in to consideration for scoring the solutions. Most often used scoring functions are namely;

Geometric/Shape Complementarity

This scoring function scores the complementarity of molecular shapes at the binding interface of the protein and the ligand. It is based on geometric features of the surface of the interacting molecules, rewarding surface contact, penalizing overlaps, and rejecting serious overlaps.

Energy Functions

Energy functions are used to evaluate how good a conformation is, as these functions generate a value for energy based on the conformation of the molecule. They provide information on what conformations of the molecule are better or worse as lower the energy value, then the better will be the conformation. The actual energy value produced by the function does not provide any useful information by itself, it's the comparison to another value that helps in analysing which conformation is better. One may conclude, that the basic property of these functions is minimization of energy of the docking complex. The terms used in energy functions for docking problem include all kinds of non-covalent interactions such as Coulombic interaction, hydrogen bonding, Van-der Waals interaction and hydrophobicity. The energy functions are mostly used as secondary energetic filters in docking for improving the ranks of nearly correct solutions obtained after an initial global search based on shape complementarity.

Scope and Limitations of Protein-Ligand Docking

Molecular Docking algorithms hold various promises for the future [1]. It can be used in *proposing potential drugs in effect enhancing drug discovery* and reducing the work of molecular biologists to large extent. One of its most important benefit is going to be able to perform *structure based drug design i.e.* proposing drugs based on the specific structure of the molecule, to which it best interacts/binds. In other words one should be able to *search a database for interacting proteins with the query.*

But, as with every technology follows the limitations, so is for *molecular docking.* Its most important limitation is incorporation of all kinds of flexibility [26] at the binding site and also the incorporation of water molecules [2] at the interface which play rather an important role during protein-protein or protein-inhibitor interactions, infact sometimes water molecules form important direct contacts at the interfaces. These two are the two important limitations of the existing docking algorithms [26]. But, possibly in future these limitations would be overcome by the development of more powerful computers and algorithms.

Chapter 2

Formulation and Algorithm Developed for the Docking Problem

2.1 Objective of the Algorithm

The objective of the algorithm presented in the thesis is to reconstruct a bound complex from unbound structures by the optimization of electrostatic interaction at the interface of the docking site. The algorithm developed here deals with protein-protein docking. The protein complexes used here are protein-inhibitor complexes. The electrostatic interaction has been calculated on the basis of point to point Coulombic interaction. Electrostatic interaction is one of the important kind of interactions occurring at protein-protein interfaces due to the presence of charged and polar side chains at the binding surfaces. It has been suggested on the basis of several studies [3] that electrostatic interaction especially play a key role in conferring specificity to the binding site and stabilization of the complex along with other kind of interactions in many cases of protein-protein association. So, the algorithm presented here, also aims at exploring the advantages and disadvantages of docking only on the basis of electrostatic interaction. For this purpose, the algorithm has been tested on 10 different protein-protein complexes and a complete analysis of the results on the basis of amino acid composition at the interface has been done. Based on the importance of electrostatic interaction, the algorithm also aims at developing a secondary screening energy filter, which can be used for improving the ranks of candidate solutions obtained after initial screening done on the basis of geometric complementarity.

2.2 Description of the Algorithm

The complete algorithm can be described in the following five parts :

2.2.1 Collection of Data from PDB and Generating an Array of Atom Co-ordinates

The PDB (Protein-Data bank) files contains the co-ordinates of the individual atoms in the protein molecule and text which describes the source of the protein, the crystallization conditions, crystal structure and refinement details. For the purpose of this algorithm only the co-ordinates of the atoms are required of both the *complex* and *unbound structures*. So, the array function extracts these co-ordinates from the PDB files. Two kinds of array function are defined, one that extracts the co-ordinates of all the protein atoms called as the *Comarray function*, while the other one that exclusively extracts the co-ordinates of the atoms that are present in the interacting residues (amino acids) present at the binding interface and has been named as *arrays function*. The atom co-ordinates are present in the ATOM field of PDB file, such that the following columns correspond to the x, y and z co-ordinates :
column 31 to 38 = x-coordinate
column 39 to 46 = y-coordinate
column 47 to 54 = z-coordinate
These co-ordinates are required for performing the translation and rotation of the inhibitor (protein as ligand) with respect to the receptor (protein).

2.2.2 Identification of the Binding-site of Protein and Ligand

The term *binding site* is used for the surface of interaction between the protein and its inhibitor (ligand) molecule, it comprises of certain number of residues (five to twenty) present at the interacting surface of both protein and the inhibitor molecule. The identification of the binding-site between the protein and inhibitor is one of the most important and critical steps for the algorithm. This has been done by using a recently developed software called CASTp which identifies pockets and cavities on the protein surface analytically, which was also reconfirmed from literature study. Once the information of residues present at the binding site is obtained, the co-ordinates of all the atoms present in these residues are extracted in the form of array, by the function developed in the first step *i.e. arrays function*. The information of binding site is required for the calculation of electrostatic interaction between the atoms of the protein and inhibitor participating at the binding site. For the partial charges on the atoms of the amino acids standard CHARMM22 charges [24] have been used. The *arrays function* contains the partial charges in addition to the co-ordinates at the binding site.

2.2.3 Translation and Rotation of the Ligand

The ligand molecule is moved towards the protein molecule by means of translation of a point (atom) (P_1) on the ligand to a point (atom) (P_o) on the protein i.e. the receptor as follows :

$$r = ||P_o - P_1|| \tag{2.1}$$

r = translation distance between the two atoms.

Now, in order to give the initial translation distance (r), two interacting atoms (present at the binding site) are chosen *i.e.* one in the each receptor and inhibitor from the already defined binding site (known from previous step). The distance between these two atoms is the distance by which the ligand has to be translated with respect to the fixed receptor. The next step is to find the orientation of the ligand with respect to the receptor which forms a minimal energy complex. In order to obtain the minimal energy orientation of ligand, an optimization routine is called.

2.2.4 Calculation and Optimization of the Electrostatic Interaction

Electrostatic interaction in the algorithm has been calculated on the basis of *point to point Coulombic interactions*. The equation of coulombs law for the force of attraction between two point charges is as follows;

$$F = kQ_1Q_2/r^2 \tag{2.2}$$

F = force experienced between the charged particles.

Q_1, Q_2 = charges of the interacting particles.

r = distance between the two charges.

here, k is the proportionality constant given by $k = 1/4\pi\epsilon$.
where, ϵ is the dielectric constant of the medium.
A function called *Enrgfun* has been developed for the calculation of electrostatic energy of interaction across the residues at the binding site between the protein and the inhibitor. The electrostatic interaction energy is calculated on the basis of point to point Coulombic interaction method as follows;

$$\Delta E = (1/D_{in} - 1/D_{out})(Q_1 Q_2/r) \tag{2.3}$$

D_{in} = dielectric constant of protein *i.e.* 2.
D_{out} = dielectric constant of outside medium (aqueous) *i.e.* 80.
Q_1, Q_2 = charge on the two interacting atoms in Coulombs.
r = distance between the two interacting atoms in Angstrom.
ΔE = Net electrostatic interaction energy between the two atoms, it is calculated in Joules.
In order to calculate the electrostatic interaction energy between the two proteins at the binding site, first the interaction energy between the interacting residues at the level of individual atoms is calculated and is then summed up to calculate the total energy of the complex. While, in order to prevent too much interpenetration of ligand into the receptor protein a penalty is applied when the distance between the two atoms becomes less than 1.8Äas follows :

$$margin = ||P_o - P_1||^2 - d^2 \tag{2.4}$$

d = 1.8Ä
P_o = a point (atom) on receptor molecule
P_1 = a point (atom) on inhibitor molecule
$if\ margin < 0$
then,

$$\Delta E = \Delta E - c(margin) \tag{2.5}$$

where, c = 2
The value of penalty was chosen after testing the algorithm on various docking systems.

An optimization function of MATLAB called *fmincon* have been used to optimize the electrostatic energy calculated from the *energy function* to obtain the minimum energy complex by rotating the ligand with respect to the receptor in all possible directions within the supplied constraints. The output of optimization routine gives the minimum energy of the complex and most importantly the orientation of the ligand *i.e.* α, β and γ in which it forms the minimal energy complex with the receptor. Now, the complete ligand molecule is in the conformation of minimum energy with respect to the receptor.

2.2.5 Calculation of the Relative Error between Docked and Original Complexes

Now, in order to validate the results, the modeled protein complex structures obtained by docking are compared to the original structure of protein-complexes. This is done by calculating distance between the atoms of the docked protein and ligand (docked complex) and similarly, the distance between the atoms of original protein-ligand complex is calculated. Now, a relative error is calculated from the difference between the distance between the atoms of docked and original structures with respect to the distance between the atoms of the original complex as follows:

$$D = D_o - D_n \tag{2.6}$$

$$N = ||D|| \tag{2.7}$$

$$E_{rel} = N/||D_o|| \tag{2.8}$$

E_{rel} = Relative error of the new docked complex with respect to the original complex.
D_n = Distance between the atoms in the original complex
D_o = Distance between the atoms in the new docked complex
D = Difference between the distances between atoms of original complex and that of the docked complex.

In order to view the docked structures the coordinates of the inhibitor in the original complex PDB file is replaced by the new co-ordinates of the inhibitor as obtained from the minimum energy conformation after optimization. This new PDB file can be viewed by any of the *molecular viewing softwares* such as RASMOL, SWISSPDBviewer and CHIME.

2.3 Summary

The aim of the algorithm developed here is to reconstruct a protein-inhibitor complex, from the unbound structures of the protein and the inhibitor by optimization of electrostatic interactions calculated by point to point interaction Coulombic method according to the equation 2.3. The receptor (protein) molecule is kept untouched, while the ligand (protein inhibitor) is translated and rotated, and energy optimization is performed by using a matlab function *fmincon*. The docked complex is compared with the original complex by calculating a relative error value for the docked complex with respect to the original complex by using the equation 2.4.

Chapter 3

Results and Discussions

3.1 Results

The results have been summarized in the tables 3.1 and 3.2, these tables list the biological systems used to test and verify the algorithm developed here, the systems have been listed along with the PDB codes of the bound structure and unbound structures and their references. The E_{rel} values *i.e* the relative error of the docked (reconstructed) complex structure with respect to the original complex structure calculated by equation 2.4 have also been listed in the table. The table 3.1 lists the E_{rel} values for docked complexes reconstructed from *unbound* structures, while table 3.2 lists the E_{rel} values for docked complexes reconstructed from *disassembled* structures.

The figures 3.1 to 3.13 plot the relative distances between atoms in the original and the docked complexes. The docked complexes in case of figures 3.1 to 3.10 have been obtained from unbound structures, while in case of figures 3.11 to 3.13 docked complexes have been obtained from disassembled structures. It can be observed from the figures that some distances between atoms in the docked complexes are comparatively much longer than that in the original complexes, it is due to the possible reasons that these regions have similar or very small potential patches and so, some other kind of interactions or contacts are occurring at these surface patches, which the present algorithm is not able to detect *i.e* take in to account such as hydrophobic contacts, Van-der Waals interaction and short range Hydrogen-bonding. While at certain regions the distance between the atoms is shorter in docked complexes as compared to that in the original complexes, it is due to certain degree of interpenetration which the algorithm has not restricted. In cases where the original and docked complexes show much variation in the plots, as expected such cases have higher corresponding value of E_{rel}.

System	PDB codes (complex;unbound)	E_{rel}	References
trypsin/leech derived trypsin inhibitor	1LDT; 1ept/1ldt	0.1894	Stubbs et *al.* 1997
trypsin/soy-bean inhibitor	1AVW; 1ept/1avu	0.2858	Song and Suh 1998
barnase/barstar	1BRS; 1bni/1bta	0.3744	Buckle et *al.* 1994
hydrolase/hydrolase inhibitor	1BVN; 1pif/2ait	0.5337	Weigand et *al.* 1995
α-chymotrypsin/HPTI	1CHO; 5cha/1ovo	0.7608	Fujinaga et *al.* 1987
acetylcholinesterase/fasciculin-II	1FSS; 2ace/1fsc	1.2547	Harel et *al.* 1995
β-trypsin/BPT1	2PTC; 2ptn/4pti	2.6305	Marquart et *al.* 1983
trypsin/Bowman-Birk inhibitor	1SMF; 2ptn/1pi2	3.4769	Huang et *al.* 1994
subtilisin/eglin-C	2SEC; 1scd/1tec	7.6844	McPhalen and James 1988
thermitase/eglin	1TEC; 1thm/2sec	7.6844	Gros et *al.* 1989

Table 3.1: List of the complexes reconstructed from unbound structures used to test and verify the algorithm developed with their respective relative error values.

System	PDB codes (complex;unbound)	E_{rel}	References
subtilisin/eglin-C	2SEC; 1scd/1tec	0.2829	McPhalen and James 1988
trypsin/Bowman-Birk inhibitor	1SMF; 2ptn/1pi2	0.4723	Huang et *al.* 1994
thermitase/eglin	1TEC; 1thm/2sec	0.8024	Gros et *al.* 1989

Table 3.2: List of the complexes reconstructed from disassembled structures used to test and verify the algorithm developed with their respective relative error values.

The following figures plot the distances between atoms of interacting residues in the original complex and the docked complex for comparison. Here, docked complex has been obtained from unbound *structures.*

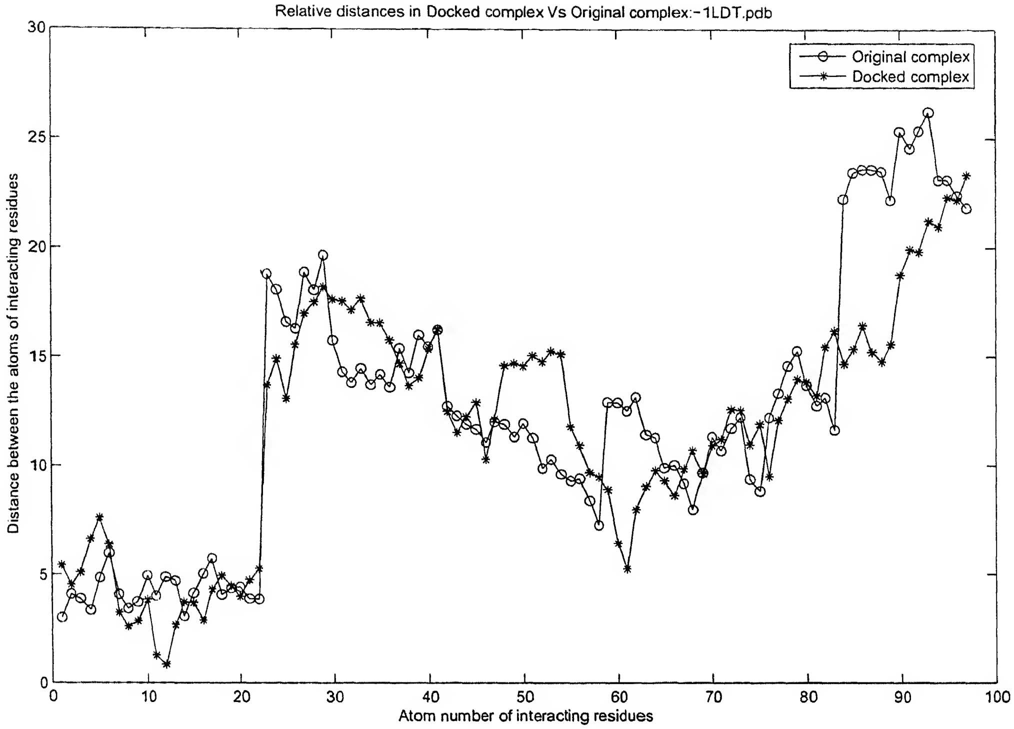


Figure 3.1: Relative distances (Å) in Docked complex Vs Original complex :- 1LDT

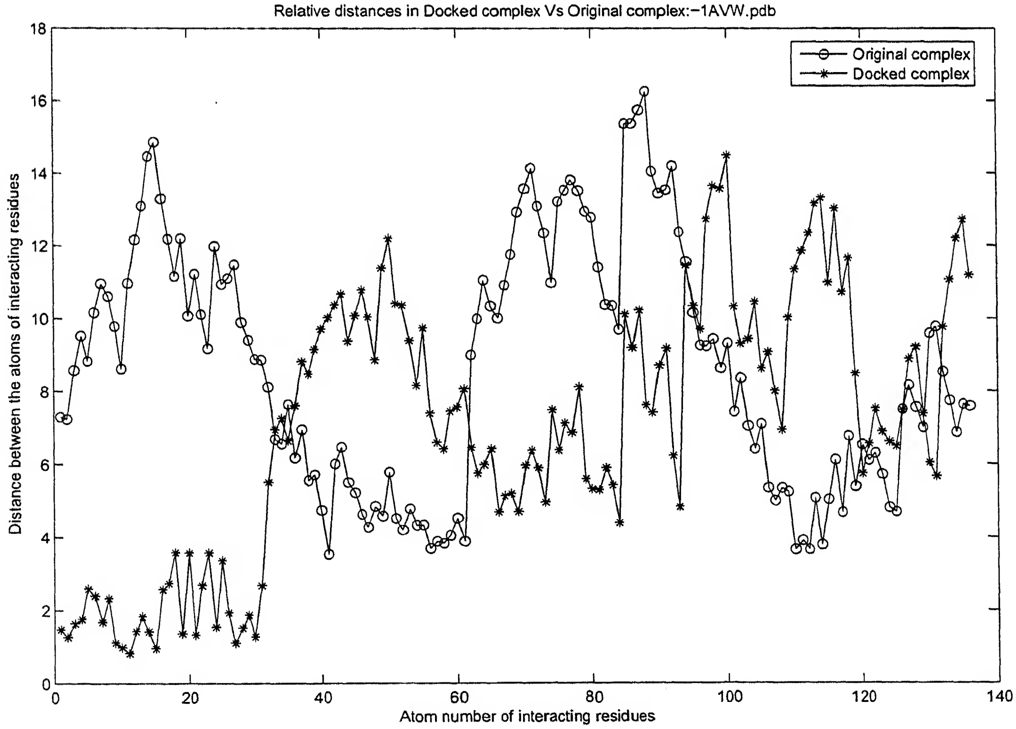


Figure 3.2: Relative distances (Å) in Docked complex Vs Original complex :- 1AVW

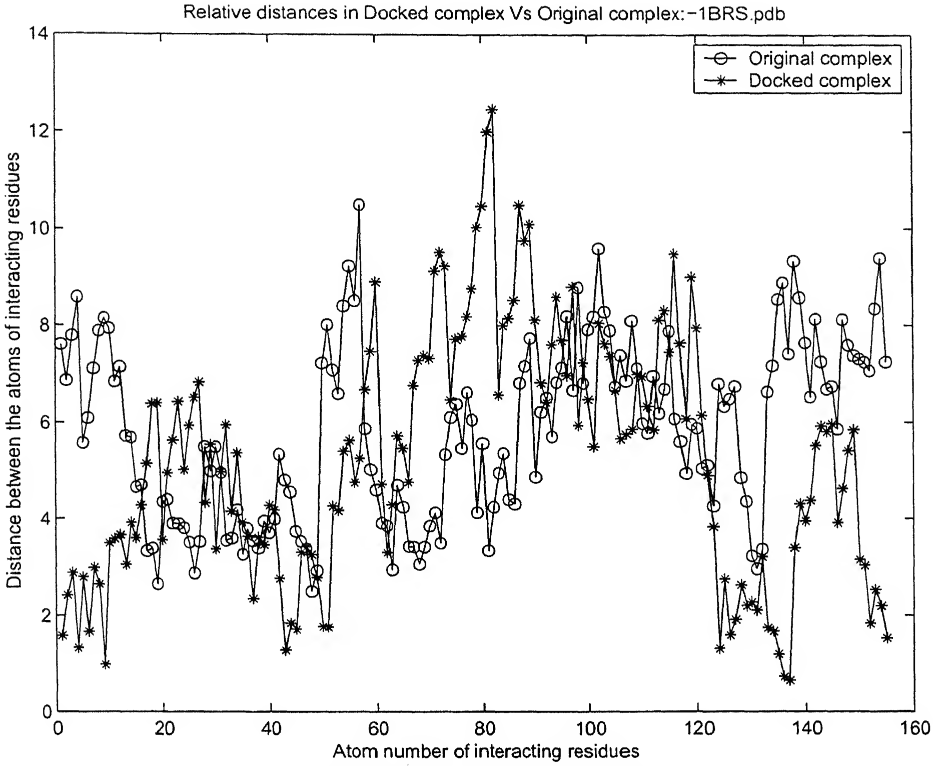


Figure 3.3: Relative distances (Å) in Docked complex Vs Original complex :- 1BRS

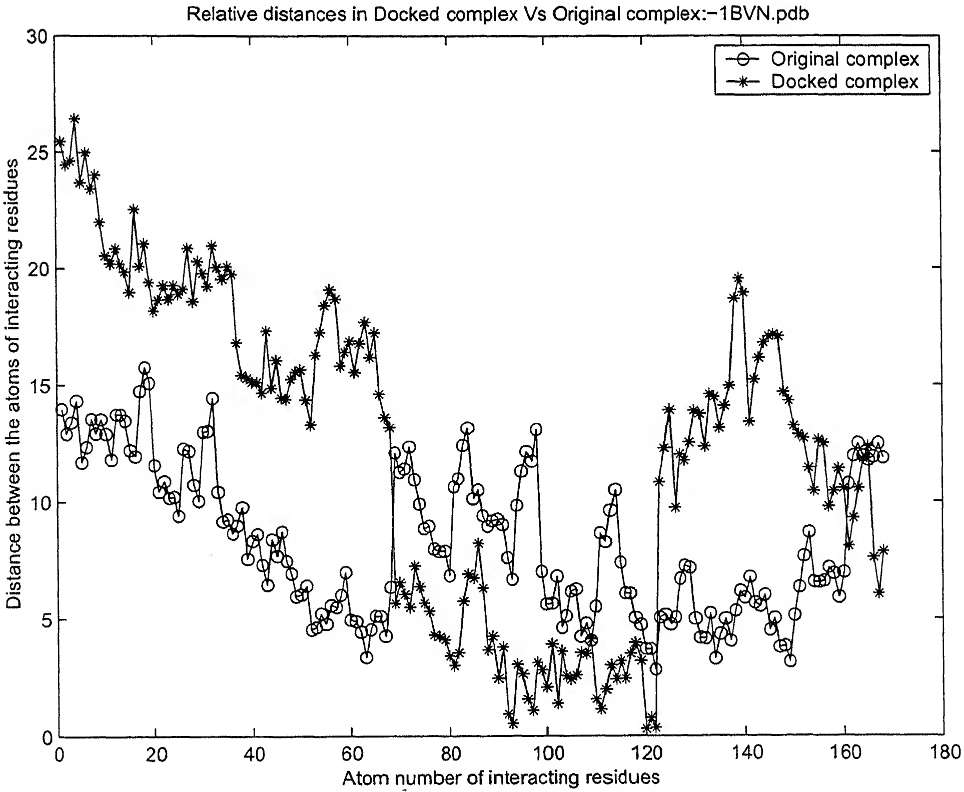


Figure 3.4: Relative distances (Å) in Docked complex Vs Original complex :- 1BVN

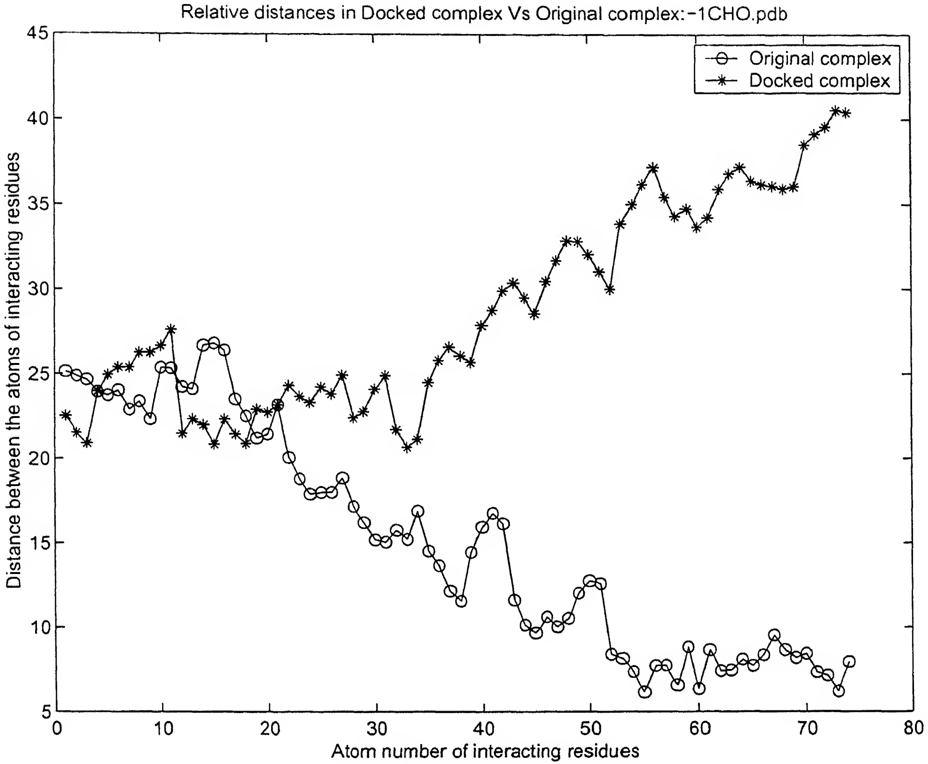


Figure 3.5: Relative distances (Å) in Docked complex Vs Original complex :- 1CHO

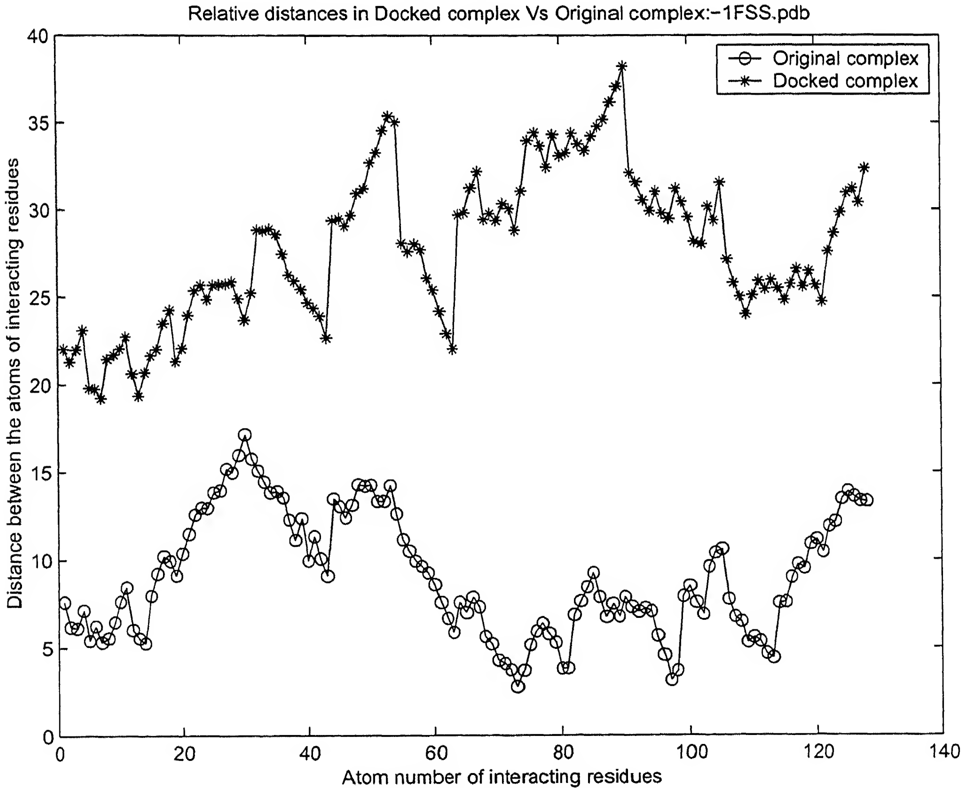


Figure 3.6: Relative distances (Å) in Docked complex Vs Original complex :- 1FSS

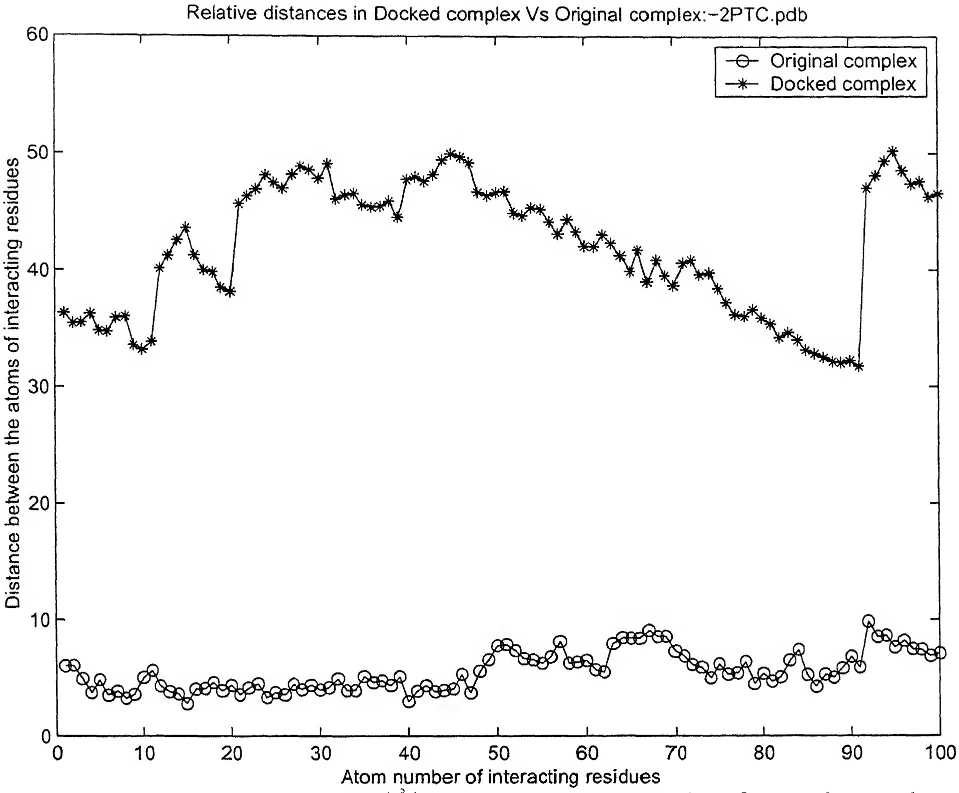


Figure 3.7: Relative distances (Å) in Docked complex Vs Original complex :- 2PTC

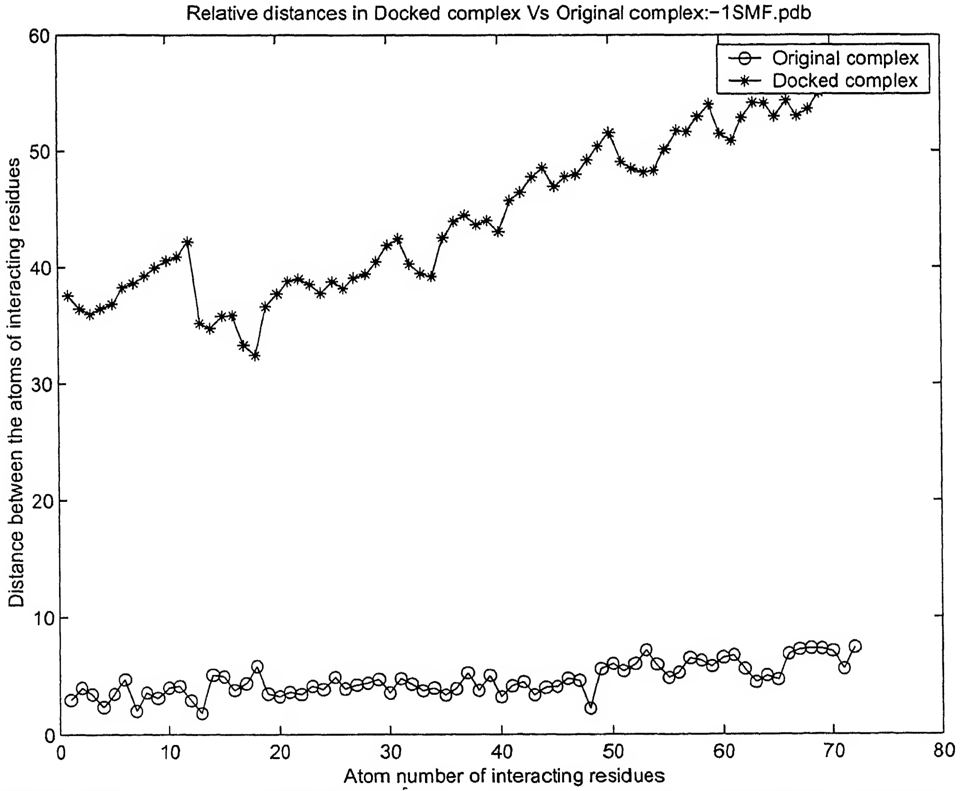


Figure 3.8: Relative distances (Å) in Docked complex Vs Original complex :- 1SMF

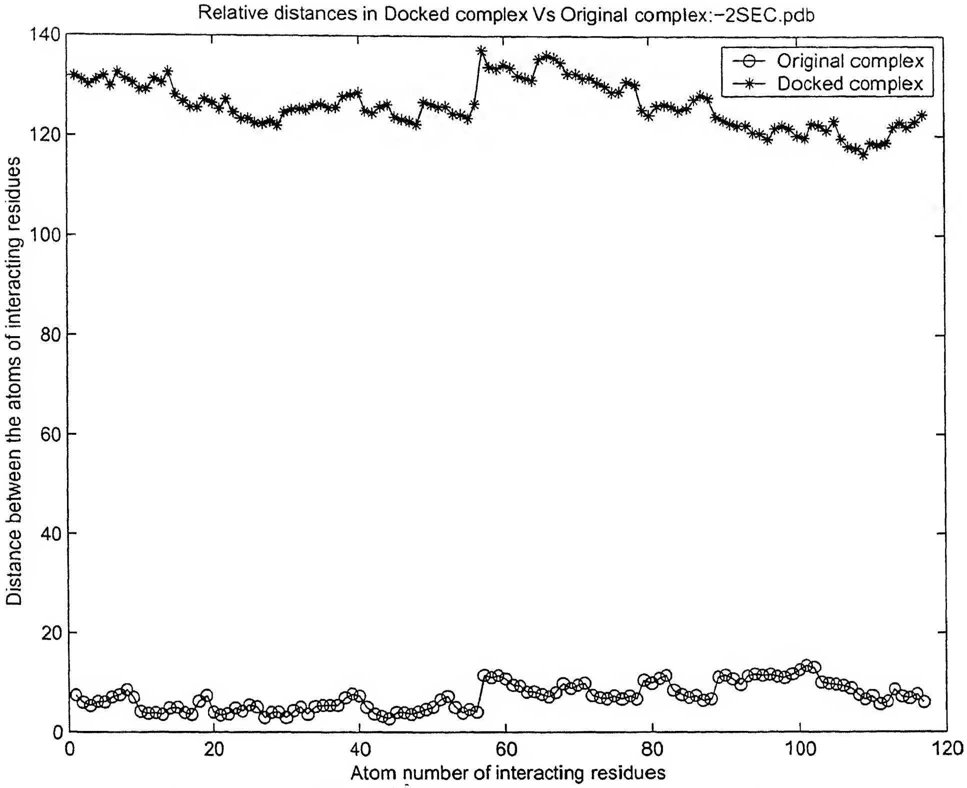


Figure 3.9: Relative distances (Å) in Docked complex Vs Original complex :- 2SEC

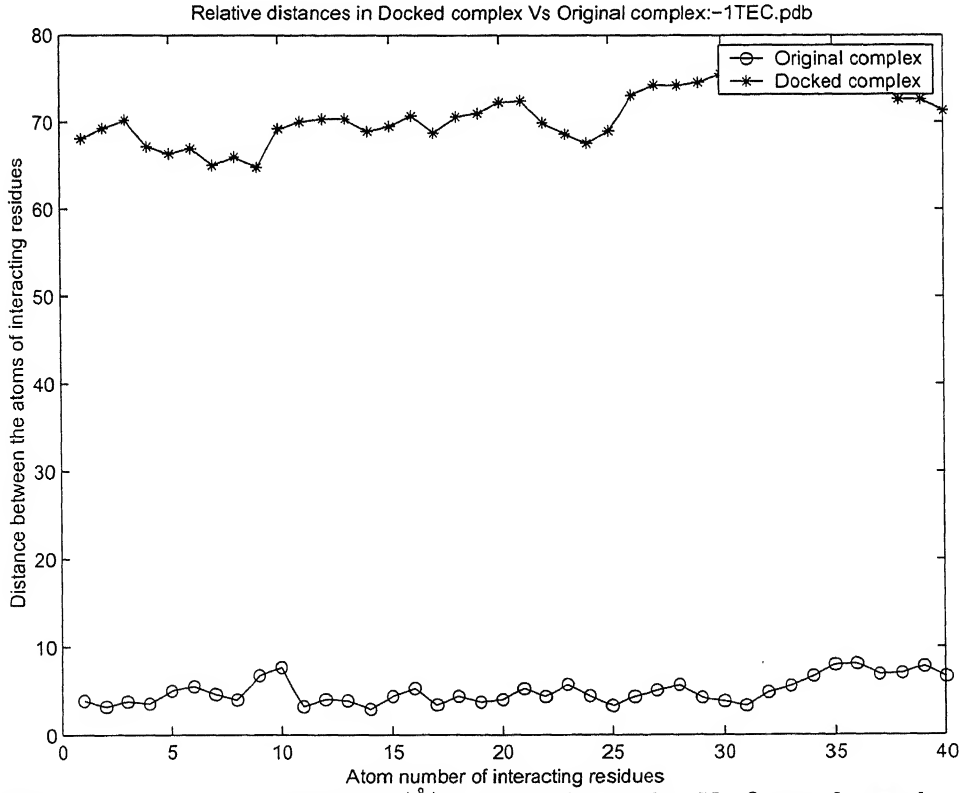


Figure 3.10: Relative distances (Å) in Docked complex Vs Original complex :- 1TEC

The following figures plot the distances between atoms of interacting residues in the original complex and the docked complex for comparison. Here, docked complex has been obtained from **dissassembled** *structures.*

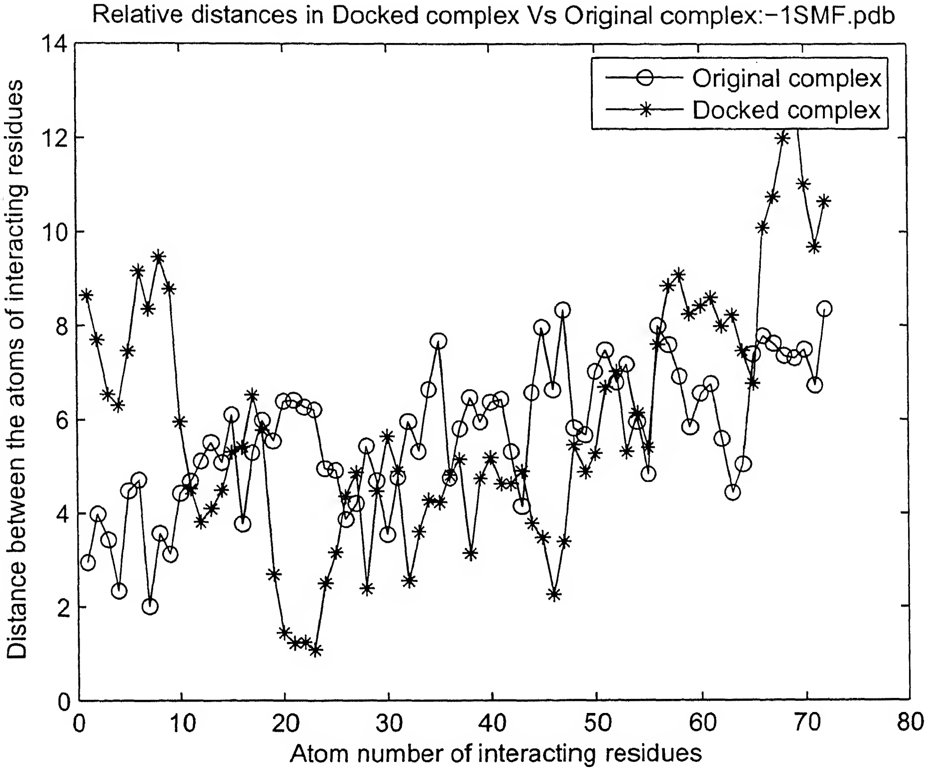


Figure 3.11: Relative distances (Å) in Docked complex Vs Original complex :- 1SMF

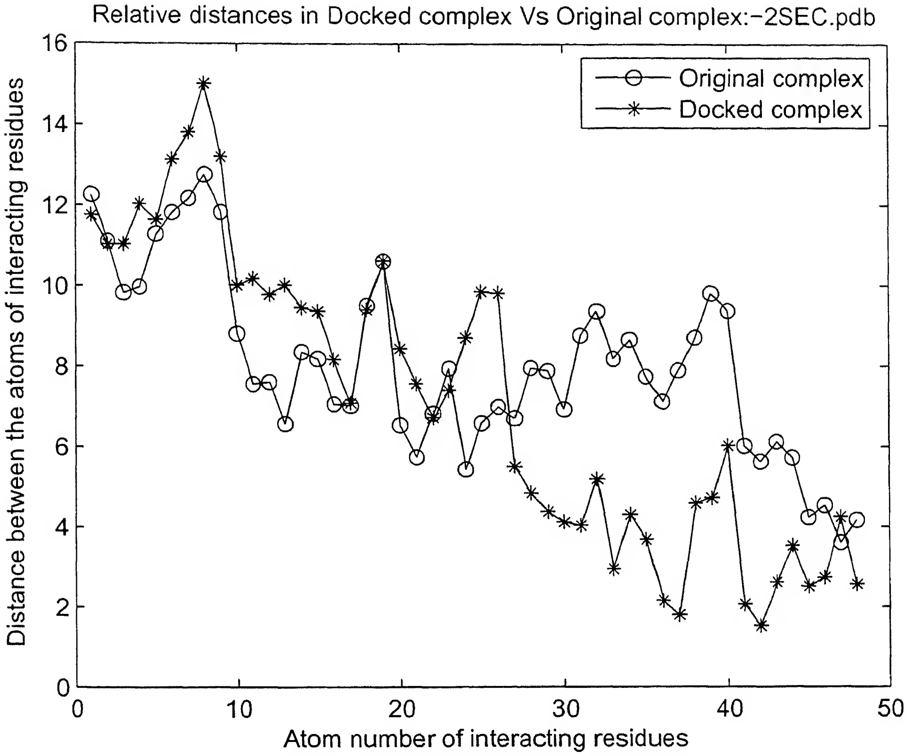


Figure 3.12: Relative distances (Å) in Docked complex Vs Original complex :- 2SEC

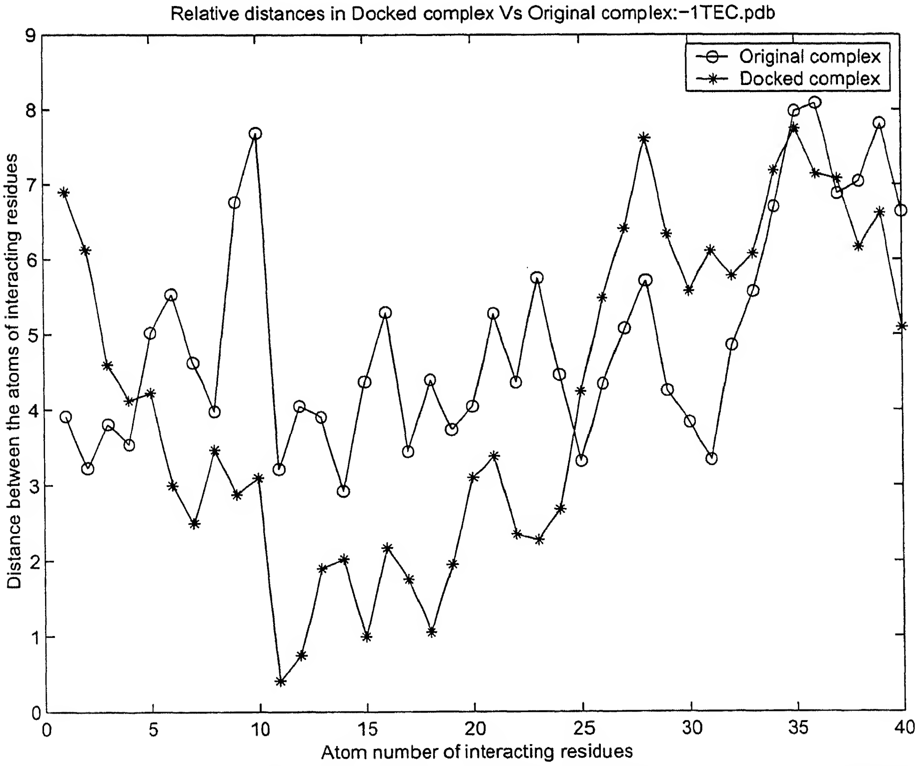


Figure 3.13: Relative distances (Å) in Docked complex Vs Original complex :- 1TEC

3.2 Discussion

The LDT complex [15] has a small ligand molecule which is 46 residues long, such small ligand molecules mostly form certain specific interactions at the binding surface, as they tend to conserve certain specific polar and charged residues at the surface. Here, both the protein and ligand molecule have many charged side chains like LYS, ARG and polar side chains like GLN and ASN at the binding surface, the presence of such polar and charged residues at the surface causes electrostatic interactions to play a key role in the complex formation. Hence, in the case of protein molecule 1LDT due to small interacting surface between the protein and the ligand, the interaction strongly depends upon the presence of these charged residues. The ligand molecule in the complex 1AVW [17], though is not of small size as is the case with 1LDT complex, but it shows presence of some charged residues (such as ASP and GLU) at the binding surface, presence of such charged residues at the surface clearly demonstrates that these residues are important for the interaction at the binding surface. The protein molecule also shows the presence of polar residues like GLN and ASN at the surface. Due to the presence of both charged and polar side chains at the binding site in 1LDT and 1AVW, electrostatic interaction helps in reconstruction of the complex with relatively low error of 18 and 28 percent, respectively.

In the complex 1BRS, both the protein and ligand molecule show the presence of charged residues LYS, ASP, GLU and ARG and some polar side chains (such as THR, TRP, TYR) at the binding surface [21], it indicates that these residues play a key role in binding. But, certain degree of hydrophobicity is also present at the interface due to the presence of non-polar residues (such as ILE, PHE, LEU, ALA, VAL) which, are equally abundant at the binding pocket. In the complex 1BVN [23] there is a large surface of interaction between the protein and ligand. Most of this surface is composed of non-polar residues (such as VAL, ALA, LEU and PRO) forming a strong hydrophobic patch at the binding site, the next most abundant residues at the surface here, are polar residues such as THR, HIS, TYR, SER, GLN and ASN, which are responsible for Hydrogen bonding at the surface of protein molecules. Thus, the presence of large interacting surface (1BVN) covered by a certain degree with a hydrophobic patch and followed by almost equal abundance of some polar and charged residues (1BRS) forming small charged potential patches at the interacting surface, it indicates that the hydrophobic patch is taking part in the stabilization of the complex apart from electrostatic interaction which is responsible for conferring specificity to the binding site. Hence, the two molecules 1BRS and 1BVN could be docked on the basis of electrostatic interaction with a relative error of 37 and 50 percent, respectively.

In the protein complex 1CHO [14], 2PTC [18] and 1FSS [16], the binding site shows the presence of hydrophobic cluster occupying 40 percent (1CHO) and 64 percent (2PTC and 1FSS) of the interacting surface [10]. This cluster consists of residues ILE, LEU, ALA, VAL, PHE, PRO, HIS, SER and TYR. The presence of such strong hydrophobic patch on the interacting surface of these complexes indicates the association of hydrophobicity as a determinant of preffered binding site between the protein and the ligand [10]. In case of protein molecule 1CHO very few charged residues (two) and some polar residues (four) are present at the surface and the hydrophobic cluster occupies 40 percent of the binding surface, so here the complex could be reconstructed by a high relative error of 76 percent. While, in case of protein molecules 2PTC and 1FSS, the hydrophobic patch occupies almost 64 percent of the binding surface, but, there is presence of *no charged residues* at the surface, thus, in these cases the complex could not be reconstructed successfully on the basis of electrostatic interaction.

This protein-inhibitor complexes 1SMF [20], 2SEC [19] and 1TEC [22] are special docking cases used here for testing the algorithm. In case of complex 1SMF, the sequence of the unbound ligand is different from that of the bound ligand present in the complex and sequence of the inhibitor has also been highly truncated *i.e.* its of size 17 residues only in the complex, while its original size is 46 residues. While, in complexes 1TEC and 2SEC, the unbound structure of the inhibitor molecule is not available, so the structure of the inhibitor in complex with some other protein has been used, but here, though both the inhibitor molecules belong to the same class, but their sequences vary with very little sequence identity. Thus, the algorithm could not reconstruct the complex successfully in these above mentioned cases from the unbound structures, so rather a different approach of docking is used for such cases termed as docking from *disassembled* structures (here, structure of the ligand is obtained from the original complex itself). This approach of docking from disassembled structures [2] has been used in these three special cases of docking *i.e.* 1SMF, 2SEC and 1TEC and the complex has been successfully reconstructed with a relative error of 47 percent, 28 percent and 68 percent, respectively. In case of complex 1SMF, the binding surface contains small charged potential patches, the surface has almost an equal distribution of polar and non-polar residues and presence of few charged residues. While, in case of 2SEC the binding pocket contains both charged and polar residues at the surface, so the complex could be reconstructed with comparatively low relative error of 28 percent. But, in case of 1TEC 78 percent of the binding surface [10] is covered by non-polar residues producing a strong hydrophobic cluster at the binding site, thus, the reconstruction of this complex on the basis of electrostatic interaction produces a high relative error.

Chapter 4

Conclusions

4.1 Summary

As observed from the results and discussions mentioned above, it can be concluded and summarized as follows :
In protein complexes where at the binding interface of the protein and the ligand are present many oppositely charged side chains and some polar side chains, such that these charges play a dominant or equivalent role (along with some other kinds of interactions) for interaction between the protein and ligand, it is possible to reconstruct the complex from the unbound structures using this algorithm with a relative error ranging from eighteen to fifty percent depending upon the role played by electrostatic interaction for the complex formation. This is observed in the initial four cases in complexes 1LDT, 1AVW, 1BRS and 1BVN and in the complexes reconstructed from disassembled structures *i.e.* 2SEC and 1SMF. In the case of protein complexes where the ligand molecule is very small in size, due to small interacting surface the interaction strongly depends on the presence of the charged side chains at the interface, this is observed in the complex 1LDT.

In those protein complexes where the binding interface has mainly nonpolar and some polar side chains, and very few or no charged side chains, such that electrostatic interaction plays a negligible or minor role in the complex formation, in such cases the relative error of reconstruction of the complex on the basis of electrostatic interaction is comparatively much high, due to the absence of strong charged potential patches. This is observed in case of protein complexes 1CHO, 2PTC and 1FSS where, at least 40 percent of the binding surface is covered by hydrophobic cluster of strong non-polar residues [10]. Thus, in such cases its not possible to dock the unbound structures on the basis of point to point Coulombic electrostatic calculations as in such cases, the presence of similar partial charges at the surface contributes to a repulsive electrostatic interaction.

In the case of protein complexes 1SMF, 2SEC and 1TEC where the *unbound* structure of the inhibitor molecule is not available due to the specific reasons mentioned in the section of discussions above, in such cases, the complex has been reconstructed from the *disassembled* structures of the protein and the ligand. This method helps in reconstruction of the complex successfully in complexes 1SMF and 2SEC with comparatively less relative error, as the binding site in these complexes has some oppositely charged potential patches. While, in case of 1TEC, the value of relative error is comparatively higher due to the presence of a strong hydrophobic patch at the binding surface which covers 78 percent of the binding surface [10] and there is absence charged residues at the binding surface.

While, performing the literature survey, working for the algorithm and testing it on various cases and based on the results and observations, apart from the above mentioned conclusions, i would also like to infer that, however, most charged groups of proteins are on the surface of the protein where they do not strongly interact with other charged groups from other proteins due to the high dielectric constant of the water solvent, but are stabilized by hydrogen bonding and polar interactions to the water and other similar side chains present on the surface of other proteins. Electrostatic interactions in water are less strong than within the protein itself (though there are few charged residues in protein interior) because of water's high dielectric constant (which results from the tendency of the large dipoles of water molecules to align with any electric field). *Electrostatic interactions can be attractive or repulsive varying from case to case depending upon the distribution of charged residues on the surface. Electrostatic interaction are responsible for conferring specificity at the binding site and their role in the formation and stabilization of the complex varies.* One more thing which has been observed is that, some PDB files have data missing (as reported in case of unbound structures of 1AVW and 1BRS) for certain number of residues or atoms present (involved in the binding site) in the *loop regions*, which are disordered, but these regions are known to frequently participate in binding site and form enzyme (protein) active sites, as they are rich in charged and polar side chains, so some information at times is missing from the PDB file, which could have been important for the energy calculations, though this does not affect the results drastically (unless there are cases where information about a complete loop participating at the binding site is missing).

The algorithm developed here, can be used in a better way on the basis of observations and conclusions obtained from the above mentioned ten cases of protein-inhibitor complexes. In order to use it in cases where the binding site information is not available and to find the approximate site of binding then in such cases, the solutions having energy of very low value of the order of 10^{-23} or even less and having a very low relative error should be considered and the structures obtained should be properly analyzed on the basis of available experimental information regarding the probable binding site (which is sometimes available in literature). Before docking one should also check the sequences of the bound and unbound ligand molecules for any differences. *This algorithm can be used for the purpose of secondary screening of the solutions obtained as possible candidates for docking on the basis of initial screening by geometric docking, in order to filter out the best out of the possible solutions*, this kind of approach has been implemented by recent geometric-electrostatic algorithms and has been found to be successful in cases where the interacting surface has large different potential patches rather than similar kind of patches or a homogeneous surface. Similarly here, *the presence of oppositely charged side chains at the interface is important for a successful docking based on point to point Coulombic interactions.*

4.2 Scope of Future Work

This algorithm uses point to point Coulombic interaction method to calculate the electrostatic interactions, where the dielectric constants for the protein is taken to be 2 and for the outside media (considering it to be aqueous) its considered 80, which is based on continuum dielectric solvation model. One can use distance dependent dielectric model and use poisson-boltzmann's equation to calculate the electrostatic potential, but that will be computationally more rigorous and expensive, in that context calculation based on the presently used method are fast. The other short range energy potentials like H-bonding, Van-der Waals interaction and Hydrophobic interactions may be included in order to calculate the complete energy of the complex.

Bibliography

[1] Inbal Halperin, Buyong Ma, HaimWolfson, and Ruth Nussinov (2002). Principles of Docking : An Overview of Search Algorithms and a Guide to Scoring Functions. *Proteins : Structure, Function, and Genetics, 47*, 409443.

[2] Talexavder Heifetz, Ephraim Katchalski-Katzir, and Miriam Eisenstein (2002). Electrostatics in protein-protein docking. *Protein Science, 11*, 571587.

[3] Felix B Sheinerman, Raquel Norel and Barry Honig (2000). Electrostatic aspects of protein-protein interactions. *Current Opinion in Structural Biology, 10*, 153159.

[4] Kay-Eberhard Gottschalk, Hani Neuvirth and Gideon Schreiber (2004). A novel method for scoring of docked protein complexes using predicted protein-protein binding sites *Protein Engineering, Design and Selection. vol. 17 no. 2*, 183-189.

[5] Fernandez-Recio, J., Totrov, M., and Abagyan, R. (2002). Screened charge electrostatic model in protein-protein docking simulations. *Pac Symp Biocomput.*, 552-63.

[6] A. Heifetz and M. Eisenstein (2003). Effect of local shape modifications of molecular surfaces on rigid-body proteinprotein docking. *Protein Engineering, Vol. 16, No. 3*, 179-185.

[7] Sharp, K. and Honig, B. (1990). Electrostatic Interactions in Macromolecules : Theory and Applications. *Ann. Rev. Biophys. Biophys. Chem 19*, 301-332.

[8] Jeffrey G. Mandell, Victoria A. Roberts, Michael E. Pique, Vladimir Kotlovyi, Julie C. Mitchell, Erik Nelson, Igor Tsigelny1 and Lynn F.

Ten Eyck1 (2001). Protein docking using continuum electrostatics and geometric fit. *Protein Engineering, Vol. 14, No. 2*, 105-113.

[9] Raquel Norel, Felix Sheinerman, Donald Petrey and Barry Honig (2001). Electrostatic contributions to proteinprotein interactions : Fast energetic filters for docking and their physical basis. *Protein Science , 10*, 2147-2161.

[10] L. Young, R.L. Jerinigan, and D.G. Covell (1994). A role for surface hydrophobicity in protein-protein recognition. *Protein Science , 3*, 717-729.

[11] David M. Lorber, Maria K. Udo and Brian K. Shoichet (2002). Proteinprotein docking with multiple residue conformations and residue substitutions. *Protein Science, 11*, 1393-1408.

[12] Henry A. Gabb, Richard M. Jackson and Michael J. E. Sternberg (1997). Modelling Protein Docking using Shape Complementarity, Electrostatics and Biochemical Information. *J. Mol. Biol. 272*, 106-120.

[13] Connolly ML. Solvent-accessible surfaces of proteins and nucleic acids. *Science, 221*, 709713.

[14] Fujinaga M, Sielecki AR, Read RJ, Ardelt W, Laskowski M Jr, James MN (1987). Crystal and molecular structures of the complex of alpha-chymotrypsin with its inhibitor turkey ovomucoid third domain at 1.8 A resolution. *J Mol Biol. 195 (2)*, 397-418.

[15] Milton T. Stubbs, Robert Morenweiser, Jrg Strzebecher, Margit Bauer, Wolfram Bode, Robert Huber, Gerd P. Piechottka, Gabriele Matschiner, Christian P. Sommerhoff, Hans Fritz and Ennes A. Auerswald (1997). The Three-dimensional Structure of Recombinant Leech-derived Tryptase Inhibitor in Complex with Trypsin. *J. of Bio. Chem. Vol. 272, No. 32*, 19931-19937.

[16] Harel M, Kleywegt GJ, Ravelli RB, Silman I, Sussman JL. (1995). Crystal structure of an acetylcholinesterase-fasciculin complex : interaction of a three-fingered toxin from snake venom with its target. *Structure. Vol.3, No. 12*, 1355-66.

[17] Song HK, Suh SW (1998). Kunitz-type soybean trypsin inhibitor revisited : refined structure of its complex with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from Erythrina caffra and tissue-type plasminogen activator. *J Mol Biol. Vol. 275, No. 2*, 347-63.

[18] M. Marquart, J. Walter, J. Deisenhofer, W. Bode and R. Huber (1983). The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors.*Acta Cryst. B39*, 480-490.

[19] McPhalen CA, James MN (1988). Structural comparison of two serine proteinase-protein inhibitor complexes : eglin-c-subtilisin Carlsberg and CI-2-subtilisin Novo.*Biochemistry. Vol.27, No.17*, 6582-98.

[20] Y. Li, Q. Huang, S. Zhang, S. Liu, C. Chi and Y. Tang (1994). Studies on an artificial trypsin inhibitor peptide derived from the mung bean trypsin inhibitor : chemical synthesis, refolding, and crystallographic analysis of its complex with trypsin. *Journal of Biochemistry, Vol. 116, Issue 1*, 18-25.

[21] Buckle AM, Schreiber G, Fersht AR. (1994). Protein-protein recognition : crystal structural analysis of a barnase-barstar complex at 2.0-A resolution. *Biochemistry. Vol.33, No.30*, 8878-89.

[22] Gros P, Fujinaga M, Dijkstra BW, Kalk KH, Hol WG. (1989). Crystallographic refinement by incorporation of molecular dynamics : thermostable serine protease thermitase complexed with eglin c. *Acta Crystallogr B. Vol. 45, Pt. 5*, 488-99.

[23] Wiegand, G. Epp, O. Huber, R. (1995). The crystal structure of porcine pancreatic alpha-amylase in complex with the microbial inhibitor Tendamistat. *J. Mol. Biol. 247*, 99-110.

[24] *Delphi Documentation.* (2000) Delphi module. CHARMM22 charges.

[25] *Insight II Documentaion.* (2000) Docking module.

[26] CAPRI : Critical Assessment of PRediction of Interactions. Documentation and Targets (http ://capri. ebi. ac. uk/)

[27] Branden, C. and Tooze, J. *Introduction to Protein Structure.* (2nd edition). Garland Publishing, New York.

[28] Creighton, T. E. *Proteins.* (2nd edition). W. H. Freeman and Co. , New York.

[29] Voet, D. and Voet, J. G. *Biochemistry.*(2nd edition). John Wiley and Sons, New York

[30] *Lehninger Principles of Biochemistry.* Nelson L. D. and Cox M. M. (3rd edition). Macmillan Worth Publishers, UK.

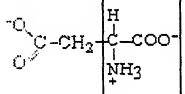
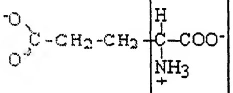
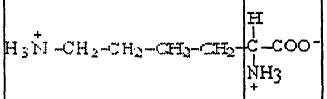
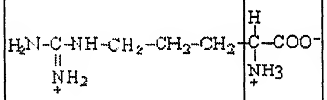
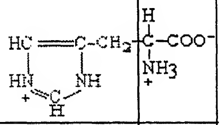
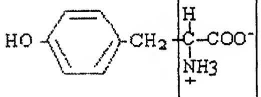
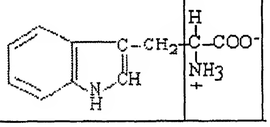
Appendix A

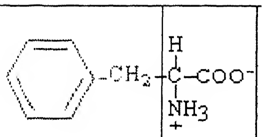
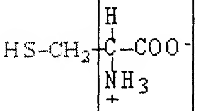
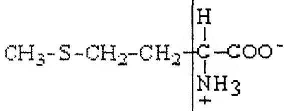
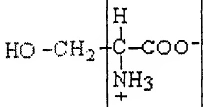
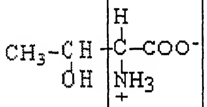
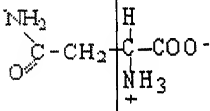
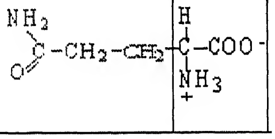
Input File Format

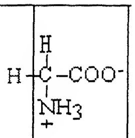
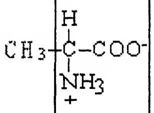
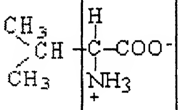
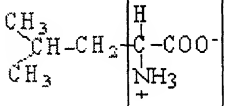
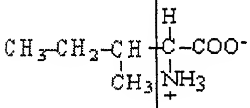
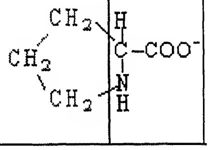
filename1='*.pdb';	enzyme file
filename2='*.pdb';	inhibitor file
filename3='*.pdb';	complex file
modelfilename='*.pdb';	new pdb filename for the model
Resvec1=[ ];	interacting residues of enzyme
ChainID1='*';	chaintype of enzyme residue
Resvec2=[ ];	interacting residues of inhibitor
ChainID2='*';	chaintype of inhibitor
Atmnum1=[ ];	atom no. of the receptor for translation
Atmnum2=[ ];	atom no. of the ligand for translation
OrgResvec1=[ ];	input('give residue nos. of original rec:')
OrgResvec2=[ ] ;	input('give residue nos. of original lig:')
OrgChainID1='*';	input('give chainid of original rec:')
OrgChainID2='*';	input('give chainid of original lig:')

Appendix B

Amino-acids Codes and Chemical Structure

Name	Symbol		R group	
	3 Lett.	1 Lett.		
Aspartate	Asp	D		
Glutamate	Glu	E		
Lysine	Lys	K		
Arginine	Arg	R		
Histidine	His	H		
Tyrosine	Tyr	Y		
Tryptophan	Trp	W		

Phenylalanine	Phe	F		
Cysteine	Cys	C		
Methionine	Met	M		
Serine	Ser	S		
Threonine	Thr	T		
Asparagine	Asn	N		
Glutamine	Gln	Q		

Glycine	Gly	G		
Alanine	Ala	A		
Valine	Val	V		
Leucine	Leu	L		
Isoleucine	Ile	I		
Proline	Pro	P		

Appendix C

Some Important URLs

1) PDB (Protein Databank)
   <http://www.rcsb.org/pdb/>
2) Molecular Viewing Softwares(RASMOL,CHIME,SWISSPDBviewer):
   <http://www.rcsb.org/pdb/help-graphics.html#rasmol_download>
3) CAPRI (Critical Assessment of PRediction of Interactions):
   <http://capri.ebi.ac.uk/>
4) CASTp:<http://cast.engr.uic.edu/cast/>
5) DELPHI documentation:
   <http://cast.engr.uic.edu/cast/>
6) INSIGHTII documentation:
   <www.chem.uh.edu/Courses/Lynch/Chem6397/Tutorials/InsightII/insight2.html>